

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

In [2]: df = pd.read_csv("Austin_weather.csv")
df

Out[2]:
```

| | Date | TempHighF | TempAvgF | TempLowF | DewPointHighF | DewPointAvgF | DewPointLowF | HumidityHighPercent | HumidityAvgPercent | HumidityLowPercent | ... | SeaLevelPressureAvgInches | SeaLe |
|------|------------|-----------|----------|----------|---------------|--------------|--------------|---------------------|--------------------|--------------------|-----|---------------------------|-------|
| 0 | 2013-12-21 | 74 | 60 | 45 | 67 | 49 | 43 | 93 | 75 | 57 | ... | 29.68 | |
| 1 | 2013-12-22 | 56 | 48 | 39 | 43 | 36 | 28 | 93 | 68 | 43 | ... | 30.13 | |
| 2 | 2013-12-23 | 58 | 45 | 32 | 31 | 27 | 23 | 76 | 52 | 27 | ... | 30.49 | |
| 3 | 2013-12-24 | 61 | 46 | 31 | 36 | 28 | 21 | 89 | 56 | 22 | ... | 30.45 | |
| 4 | 2013-12-25 | 58 | 50 | 41 | 44 | 40 | 36 | 86 | 71 | 56 | ... | 30.33 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1314 | 2017-07-27 | 103 | 89 | 75 | 71 | 67 | 61 | 82 | 54 | 25 | ... | 29.97 | |
| 1315 | 2017-07-28 | 105 | 91 | 76 | 71 | 64 | 55 | 87 | 54 | 20 | ... | 29.9 | |
| 1316 | 2017-07-29 | 107 | 92 | 77 | 72 | 64 | 55 | 82 | 51 | 19 | ... | 29.86 | |
| 1317 | 2017-07-30 | 106 | 93 | 79 | 70 | 68 | 63 | 69 | 48 | 27 | ... | 29.91 | |
| 1318 | 2017-07-31 | 99 | 88 | 77 | 66 | 61 | 54 | 64 | 43 | 22 | ... | 29.97 | |

1319 rows x 21 columns

```
In [13]: df["PrecipitationSumInches"]

Out[13]:
```

| | |
|------|------|
| 0 | 0.46 |
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | T |
| ... | ... |
| 1314 | 0 |
| 1315 | 0 |
| 1316 | 0 |
| 1317 | 0 |
| 1318 | 0 |

Name: PrecipitationSumInches, Length: 1319, dtype: object

```
In [14]: df = df.drop(['Events', 'Date', 'SeaLevelPressureHighInches', 'SeaLevelPressureLowInches'], axis = 1)
df

Out[14]:
```

| | TempHighF | TempAvgF | TempLowF | DewPointHighF | DewPointAvgF | DewPointLowF | HumidityHighPercent | HumidityAvgPercent | HumidityLowPercent | SeaLevelPressureAvgInches | VisibilityHighMiles |
|------|-----------|----------|----------|---------------|--------------|--------------|---------------------|--------------------|--------------------|---------------------------|---------------------|
| 0 | 74 | 60 | 45 | 67 | 49 | 43 | 93 | 75 | 57 | 29.68 | 10 |
| 1 | 56 | 48 | 39 | 43 | 36 | 28 | 93 | 68 | 43 | 30.13 | 10 |
| 2 | 58 | 45 | 32 | 31 | 27 | 23 | 76 | 52 | 27 | 30.49 | 10 |
| 3 | 61 | 46 | 31 | 36 | 28 | 21 | 89 | 56 | 22 | 30.45 | 10 |
| 4 | 58 | 50 | 41 | 44 | 40 | 36 | 86 | 71 | 56 | 30.33 | 10 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1314 | 103 | 89 | 75 | 71 | 67 | 61 | 82 | 54 | 25 | 29.97 | 10 |
| 1315 | 105 | 91 | 76 | 71 | 64 | 55 | 87 | 54 | 20 | 29.9 | 10 |
| 1316 | 107 | 92 | 77 | 72 | 64 | 55 | 82 | 51 | 19 | 29.86 | 10 |
| 1317 | 106 | 93 | 79 | 70 | 68 | 63 | 69 | 48 | 27 | 29.91 | 10 |
| 1318 | 99 | 88 | 77 | 66 | 61 | 54 | 64 | 43 | 22 | 29.97 | 10 |

1319 rows x 17 columns

```
In [15]: # some values have 'T' which denotes trace rainfall
# we need to replace all occurrences of T with 0
# the data also contains '-' which indicates no
# or WZ. This means that data is not available
df = df.replace(["T", "-", "WZ"], 0.0)
df["PrecipitationSumInches"]

Out[15]:
```

| | |
|------|------|
| 0 | 0.46 |
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| ... | ... |
| 1314 | 0 |
| 1315 | 0 |
| 1316 | 0 |
| 1317 | 0 |
| 1318 | 0 |

Name: PrecipitationSumInches, Length: 1319, dtype: object

```
In [16]: df.to_csv("austin_final.csv")

In [17]: df_cleaned = pd.read_csv("austin_final.csv")
df_cleaned

Out[17]:
```

| | Unnamed: 0 | TempHighF | TempAvgF | TempLowF | DewPointHighF | DewPointAvgF | DewPointLowF | HumidityHighPercent | HumidityAvgPercent | HumidityLowPercent | SeaLevelPressureAvgInches | Visibi |
|------|------------|-----------|----------|----------|---------------|--------------|--------------|---------------------|--------------------|--------------------|---------------------------|--------|
| 0 | 0 | 74 | 60 | 45 | 67.0 | 49.0 | 43.0 | 93.0 | 75.0 | 57.0 | 29.68 | |
| 1 | 1 | 56 | 48 | 39 | 43.0 | 36.0 | 28.0 | 93.0 | 68.0 | 43.0 | 30.13 | |
| 2 | 2 | 58 | 45 | 32 | 31.0 | 27.0 | 23.0 | 76.0 | 52.0 | 27.0 | 30.49 | |
| 3 | 3 | 61 | 46 | 31 | 36.0 | 28.0 | 21.0 | 89.0 | 56.0 | 22.0 | 30.45 | |
| 4 | 4 | 58 | 50 | 41 | 44.0 | 40.0 | 36.0 | 86.0 | 71.0 | 56.0 | 30.33 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1314 | 1314 | 103 | 89 | 75 | 71.0 | 67.0 | 61.0 | 82.0 | 54.0 | 25.0 | 29.97 | |
| 1315 | 1315 | 105 | 91 | 76 | 71.0 | 64.0 | 55.0 | 87.0 | 54.0 | 20.0 | 29.90 | |
| 1316 | 1316 | 107 | 92 | 77 | 72.0 | 64.0 | 55.0 | 82.0 | 51.0 | 19.0 | 29.86 | |
| 1317 | 1317 | 106 | 93 | 79 | 70.0 | 68.0 | 63.0 | 69.0 | 48.0 | 27.0 | 29.91 | |
| 1318 | 1318 | 99 | 88 | 77 | 66.0 | 61.0 | 54.0 | 64.0 | 43.0 | 22.0 | 29.97 | |

1319 rows x 18 columns

```
In [ ]:

In [ ]:

In [18]: x = df_cleaned.drop(['PrecipitationSumInches'], axis = 1)
y = df_cleaned.iloc[:, -1]
y

Out[18]:
```

| | |
|------|------|
| 0 | 0.46 |
| 1 | 0.00 |
| 2 | 0.00 |
| 3 | 0.00 |
| 4 | 0.00 |
| ... | ... |
| 1314 | 0.00 |
| 1315 | 0.00 |
| 1316 | 0.00 |
| 1317 | 0.00 |
| 1318 | 0.00 |

Name: PrecipitationSumInches, Length: 1319, dtype: float64

```
In [33]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)
x_scaled

Out[33]: array([[ -1.73973815, -0.46492853, -0.75801089, ..., 1.9621461 ,
-0.47810712, 1.62634133],
[ -1.72811184, -1.68436437, -1.61267654, ..., 0.80599318,
0.47738272, 0.61718235],
[ -1.72548553, 1.5488715 , -1.82634295, ..., -1.50634267,
-0.95585205, -1.55632879],
...,
[ 1.72548553, 1.77070384, 1.52109751, ..., -0.35017975,
-0.47810712, -0.72036297],
[ 1.72811184, 1.7029574 , 1.59231964, ..., -0.06113902,
-0.47810712, -0.21879347],
[ 1.73973815, 1.22873235, 1.23620896, ..., -0.35017975,
-0.47810712, -0.21878347]])

In [34]: y_ = y.values.reshape(-1,1)
y_

Out[34]: array([[0.46],
[0. ],
[0. ],
...,
[0. ],
[0. ],
[0. ]])

In [35]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x_scaled, y_, test_size=0.2)
print(x_train.shape, x_test.shape)

(1055, 17) (264, 17)

In [36]: from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
y_pred

Out[36]: array([[ 0.00885603],
[ 0.33518663],
[ 0.23535087],
[-0.0441561 ],
[ 0.18065995],
[-0.95368249],
[ 0.19182322],
[ 0.02612817],
[ 0.22480909],
[-0.93550141],
[ 0.08266431],
[ 0.24321132],
[-0.95360531],
[ 0.13788462],
[-0.05954595],
[ 0.43740544],
[ 0.26730808],
[ 0.06826243],
[ 0.24972021],
[ 0.25727833],
[ 0.02317240],
[ 0.14457464],
[ 0.40919157],
[ 0.07729675],
[-0.06912706],
[-0.95315507],
[-0.12129708],
[-0.04079321],
[ 0.27940167],
[ 0.95060511],
[ 0.06840337],
[ 0.20456823],
[-0.00645864],
[ 0.123392 ],
[-0.04662583],
[ 0.18994592],
[ 0.12549041],
[ 0.19115703],
[ 0.05121745],
[ 0.10024180],
[ 0.03152588],
[ 0.29123983],
[ 0.1270099 ],
[ 0.0190535 ],
[-0.09788636],
[-0.08113127],
[ 0.71891505],
[-0.01864376],
[ 0.83468348],
[-0.10060467],
[ 0.15947883],
[ 0.17835149],
[ 0.26277604],
[ 0.03506092],
[ 0.07971165],
[ 0.00644935],
[ 0.30924810],
[ 0.06219707],
[-0.10277658],
[ 0.61333143],
[ 0.246573 ],
[-0.07367397],
[-0.00323735],
[ 0.05063019],
[-0.12356519],
[-0.06519874],
[-0.10075645],
[-0.08106257],
[ 0.25804787],
[ 0.02470462],
[ 0.1250106 ],
[ 0.20798515],
[-0.03610698],
[-0.06074527],
[-0.11403212],
[-0.0567159 ],
[ 0.26970462],
[ 0.08853019],
[ 0.18572426],
[ 0.04701095],
[ 0.4306069 ],
[ 0.13143141],
[ 0.0708346 ],
[ 0.0522197 ],
[ 0.16811519],
[ 0.63134681],
[ 0.2196047 ],
[ 0.73545664],
[-0.14450141],
[ 1.03302960],
[ 0.01550075],
[ 0.47154209],
[ 0.03163281],
[-0.10132421],
[ 0.02639645],
[ 0.37915326],
[ 0.02369357],
[ 0.02670571],
[ 0.05490038],
[ 0.02469636],
[ 0.0501246 ],
[ 0.09650066],
[ 0.50903177],
[ 0.06393433],
[ 0.04510819],
[ 0.54977219],
[ 0.13960669],
[-0.05457458],
[ 0.11510803],
[-0.1625763 ],
[ 0.11493136],
[ 0.05327525],
[ 0.1957886 ],
[ 0.09358912],
[ 0.09742524],
[ 0.43073537],
[ 0.08442154],
[ 0.1586923 ],
[ 0.13054053],
[ 0.12249434],
[-0.08211762],
[ 0.27608973],
[ 0.08052326],
[ 0.00315971],
[ 0.02670841],
[ 0.05087051],
[ 0.11708366],
[ 0.09592454],
[ 0.44811907],
[-0.00240966],
[ 0.1821302 ],
[ 0.03174730],
[ 0.01201673],
[-0.00234237],
[ 0.07652645],
[-0.01197071],
[ 0.2789108 ],
[-0.09379528],
[ 0.03116074],
[-0.073827 ],
[-0.0015557 ],
[ 0.05034847],
[-0.122138 ],
[-0.00346215],
[ 0.00260884],
[ 0.17156304],
[ 0.5116628 ],
[ 0.45430740],
[ 0.09105402],
[ 0.44867902],
[-0.04900095],
[ 0.00629403],
[ 0.04006417],
[ 0.00355244],
[ 0.04466343],
[ 0.32112266],
[ 0.29314659],
[-0.07060157],
[ 0.05959212],
[-0.00489102],
[ 0.34090115],
[ 0.01301637],
[ 0.0210512 ],
[-0.11250621],
[ 0.3016449 ],
[ 0.03340273],
[-0.10599275],
[ 0.20038577],
[ 0.09070554],
[ 0.06072887],
[ 0.00490904],
[ 0.52908585],
[ 0.00329604],
[ 0.15977906],
[ 0.11014719],
[ 0.11034627],
[ 0.05214701],
[-0.02143204],
[ 0.14457555],
[ 0.51704590],
[ 0.32503478],
[ 0.21730577],
[ 0.15015170],
[-0.10138801],
[-0.05124771],
[-0.05087419],
[ 0.05507515],
[ 0.04107814],
[ 0.12850604],
[ 0.11037646],
[-0.12131685],
[ 0.36301858],
[ 0.05051532],
[-0.03367407],
[ 0.11701691],
[ 0.04706071],
[ 0.07355936],
[ 0.10399622],
[ 0.12019964],
[ 0.03413543],
[ 0.65395544],
[ 0.26387006],
[ 0.25542670],
[-0.14810435],
[ 0.16230734],
[ 0.14705035],
[ 0.45225492],
[ 0.36477599],
[ 0.20680098],
[ 0.09005040],
[ 0.0124778 ],
[-0.00973933],
[ 0.01559103],
[ 0.37677508],
[ 0.03326145],
[-0.00730799],
[ 0.00813072],
[ 0.01294518],
[ 0.15424916],
[ 0.02019302],
[-0.11180316],
[ 0.42876883],
[ 0.01683204],
[-0.12403931],
[ 0.12854164],
[ 0.31425781],
[ 0.3305111 ],
[-0.136634 ],
[ 0.21407388],
[ 0.04303771],
[ 0.22344311],
[-0.1331359 ],
[ 0.07079753],
[ 0.31753901],
[-0.00207384],
[ 0.02237397],
[-0.00340408],
[ 0.1794313 ],
[ 0.37061018],
[ 0.17914423],
[-0.00920022],
[ 0.47260561],
[ 0.20462442],
[ 0.57183865],
[ 0.33251211],
[ 0.02480957],
[ 0.03089040],
[ 0.50932891],
[-0.11652724],
[ 0.17095717],
[ 0.70655710],
[ 0.11802322],
[ 0.07840638],
[ 0.05006358],
[ 0.04142998],
[ 0.00764492],
[ 0.03067109],
[ 0.00377011],
[ 0.14905080],
[ 0.23310004],
[-0.05954283],
[-0.24749463],
[ 0.31257745],
[ 0.25204047],
[-0.16179636]])

In [60]: day_index = 200
days = [i for i in range(y_.size)]
len(days)

Out[60]: 1319

In [61]: print("The precipitation trend graph: ")
plt.scatter(days, y_, color = 'g')
plt.scatter(days[day_index], y_[day_index], color = 'r')
plt.title("Precipitation level")
plt.xlabel("Days")
plt.ylabel("Precipitation in inches")
plt.show()

The precipitation trend graph:


```